



Modern Data Warehousing without the Burden of ETL

By Dave Wells

May 2018

Research Sponsored by

incorta.

About the Author



Dave Wells is an advisory consultant, educator, and research analyst dedicated to building meaningful connections throughout the path from data to business impact. As an educator he has written dozens of courses and taught hundreds of classes about data warehousing, data modeling, data architecture, and business intelligence for professional organizations such as Dataversity, eLearningCurve, and TDWI.

Dave works at the intersection of information and business, driving value through analytics, business intelligence, and innovation. More than 40 years of information management experience combined with over 10 years of business management create a unique perspective about the connections among business, information, data, and technology. Knowledge sharing and skills building are Dave's passions, carried out through consulting, speaking, teaching, and writing.

About Eckerson Group

Eckerson Group is a research and consulting firm of veteran practitioners who help business analytics leaders use data and technology to drive better insights and actions. Its researchers and consultants each have more than 20 years of experience in the field and are uniquely qualified to help business and technical leaders optimize their investments in business intelligence, analytics, big data management, and the internet of things.



To see our research and learn about our consulting services, go to www.eckerson.com.

Abstract

Most organizations today operate and support multiple data warehouses with a typical company having two to five legacy data warehouses. Business people and processes depend on these data warehouses, but their operations and support challenges are growing. Past practices for data warehousing struggle to meet today's needs. Legacy data warehouses don't scale easily, are performance challenged, and are constrained by relational models for structured data. They take time to build, deploy, and change.

The time has come to rethink and remake data warehousing. Despite the challenges, we can't simply decommission legacy data warehouses and rely solely on a data lake. We must have a way to sustain the benefits of data warehousing—integrated, time-variant enterprise data that is easy to access and understand—without the burden of time-consuming and complex data modeling and ETL processing.

Incorta's innovative approach of Direct Data Mapping meets the challenges of modern data warehousing with the right techniques and tools to deliver high-speed data integration and aggregation with all of the advantages of a semantic layer, and without the burden of a 1990s-style data warehouse.

This Incorta product profile is a companion to an Eckerson Group report titled [Modern Data Warehousing: Analytics without the Modeling](#) (August 2017). This report is also a companion to the webinar, [Modern Data Warehousing without the Burden of ETL](#).

The Legacy Data Warehouse Problem

Remember the data warehousing promises of the past—data organized for ease of access and understanding, data at the speed of business, a single version of the truth, etc. Today nearly every organization operates at least one data warehouse and most have two or more. Yet the goals of fast, easy, and single source remain elusive. Data warehouses that were once sources of pride are now seen as troublesome legacy systems.

Past practices for data warehousing struggle to meet today's needs in the age of big data and fast analytics. Legacy data warehouses don't scale easily, are performance challenged, and are constrained by relational models for structured data. Perhaps more important, they are slow—slow to build, slow to deploy, slow to change, and slow to deliver. Time-consuming projects that deliver high-latency data simply fall short of service level requirements for modern data management.

How do we get the benefits of data warehousing without the challenges of managing a data warehouse?

The time has come to rethink and remake data warehousing. Despite their challenges, we can't simply decommission data warehouses and rely solely on a data lake. The primary purposes of a data warehouse are integration and reconciliation of enterprise data and retention of time-variant snapshots of business history. Failure to integrate and organize data limits the ability to publish the kinds of information that people use routinely in their jobs. The publish-and-subscribe model works and people still want information to be published for them. Lack of time-variant history with uniform time intervals inhibits the ability to perform time-series analysis and to understand trends in the business. The big question when seeking to modernize legacy data warehousing is how to realize the benefits of data warehousing without facing the challenges of managing a data warehouse.

The Data Warehouse Alternative

Direct Data Mapping™

The magic of the Incorta solution is innovative Direct Data Mapping technology that shifts the data integration workload from data ingestion to data consumption. Without the requirement to reshape the data—integrate, aggregate, dimensionalize, etc.—data ingestion is simpler and faster. Direct Data Mapping connects the dots at query time. It delivers near-real-time reporting and analytics directly from original transactional data such as legacy systems, ERP systems, and SaaS applications.

One-to-one mapping starts the mapping process—one source file or table is mapped directly to one Incorta table. Next, mapping to support joins is added. Join capabilities that exist in data sources are automatically mapped. Relationships not present in source data are added to enrich Incorta's ability to join, integrate, and aggregate the data. New relationships can be found by manual detection and by reading existing metadata. The Incorta roadmap includes a feature that will look at existing SQL statements to profile join information and automatically build relationships.

Direct Data Mapping™ works with transactional data and also supports opportunities to enrich the analytical experience by combining it with data from a variety of sources. Plug-and-play data connectors support inclusion of flat files, Excel, CSV/JSON, Hive, and JDBC data sources along with relational databases, Kafka streaming data, and SaaS applications such as Salesforce.com, Marketo, and ServiceNow.

Free from Modeling and ETL

Conventional data warehouses are slow, complex, and difficult to manage. Developing data models to impose dimensional structure on the data warehouse takes time and, perhaps more importantly,

inhibits adaptability to new and changing business requirements. Building, operating, and maintaining the ETL processes that restructure the data is complex, labor intensive, and costly. Batch ETL processing is also a primary driver of data latency.

The Incorta alternative uses Apache Parquet in place of a data warehouse. Parquet is a compressed and highly efficient columnar database that meets the need to capture and store enterprise data and to collect historical data that is essential for time-series analysis. Parquet offers many advantages for data management, including reduced file size through compression, performance gains with parallel processing efficiencies, and the abilities to store complex nested data structures and break away from relational model constraints. Building on the Parquet foundation, Incorta provides exceptionally fast data ingestion from a variety of sources in near real time.

The challenges of data warehouse management are eliminated without loss of data or information.

Incorta moves data from sources directly into Parquet without aggregation, normalization, dimensionalizing, reformatting, or otherwise transforming the data. This approach eliminates data staging tables, star schema design, source-to-target mapping, ETL scripting, surrogate generation, key mapping, and other complexities of traditional ETL.

As a result, the time, labor, and cost of ETL development simply disappear. Data latency that is inherent in batch ETL processing is reduced or eliminated. Difficult data structures from legacy sources cease to be a problem with Parquet flexibility. Parquet readily adapts to new and changing data sources. Tracing data lineage is substantially simplified when complex transformations are removed. In short, the challenges of data warehouse management are eliminated with no loss of data or information.

High-Speed Queries

With Direct Data Mapping Incorta performs incredibly fast queries, with typical response times reduced from hours to seconds, even with complex joins and aggregations that combine data from hundreds of tables and billions of rows.

The traditional role of a data warehouse is unnecessary. There is no need to pre-aggregate and pre-join data before analysis.

With Incorta the traditional role of a data warehouse—advance data integration and optimization for query and analysis—is unnecessary. High-speed query is possible without the need for premature decisions about aggregation and hierarchical drill path organization. Unlike star schema where the number of dimensions directly affects query performance, data mapping supports unlimited dimensionality without loss of performance. With Incorta, each query may dimensionalize, aggregate, pivot, search, and filter uniquely without loss of performance. Individual analysis and reporting use cases are no longer constrained by a one-size-fits-all data model.

Seamless Analytics

Data visualization and analytics features are built into Incorta for a seamless user experience. Starting with data discovery, Incorta offers a Google-like search feature that helps analysts find the best-fit data for every analytics use case. Business users can use Incorta's built-in data visualization features, and they can also work with familiar and comfortable tools including Excel, Tableau, Power BI, and Qlik while enjoying substantial performance gains. Data analysis and visualization occur without compromising data security because all application security constraints carry forward to Incorta and are applied at data access time. Advanced analytics capabilities include predictive modeling and machine learning, which use native integration of Parquet and Spark as components of the Hadoop ecosystem.

Data mapping that spans the continuum from data sources to business semantics is a valuable feature that enables users to work with familiar concepts and terminology without concern for data location, structure, or technical details. Secure, business-friendly, and seamless analytics provides a fast path to meaningful business insights.

Pain-Free Deployment

Incorta readily adapts to your deployment preferences whether on premises or in the cloud. On premises deployment uses low-cost commodity hardware. Cloud deployment is accelerated with provided templates for AWS, Microsoft Azure, and Google Cloud. Regardless of deployment choice, the elastic architecture is designed to adapt resource allocation to match dynamic workload demands.

Modernizing Your Legacy Data Warehouses

Many organizations are considering cloud migration as a path to data warehouse modernization. The cloud has many benefits, but the migration process is labor-intensive and time consuming. Moving data is only the beginning of cloud data warehousing. You must migrate schema, data, metadata, ETL processing, users, and applications. Even with a cloud deployment Direct Data Mapping eliminates much migration complexity.

We need data warehousing without the burden of the traditional data warehouse.

Whether you have one data warehouse or many, the need to modernize is a pressing reality. Most organizations today operate and support multiple data warehouses. A typical company has two to five legacy data warehouses as a result of mergers, acquisitions, independent departmental initiatives, and other causes. Multiplicity of data warehouses creates user confusion and uncertainty. With Incorta you can map the tables of all legacy warehouses so they appear to the business as a single cohesive data resource. Immediate benefits are increased confidence when working with data, search capability across all legacy warehouse data, and visible gains in query performance.

To expand the value and impact of modernization, retire old and challenging batch ETL processes. Begin with the most painful processes and eliminate them by mapping the original sources in Incorta. Work incrementally to eliminate the ETL processes that are the greatest management challenges and those where data latency has the most severe business impacts.

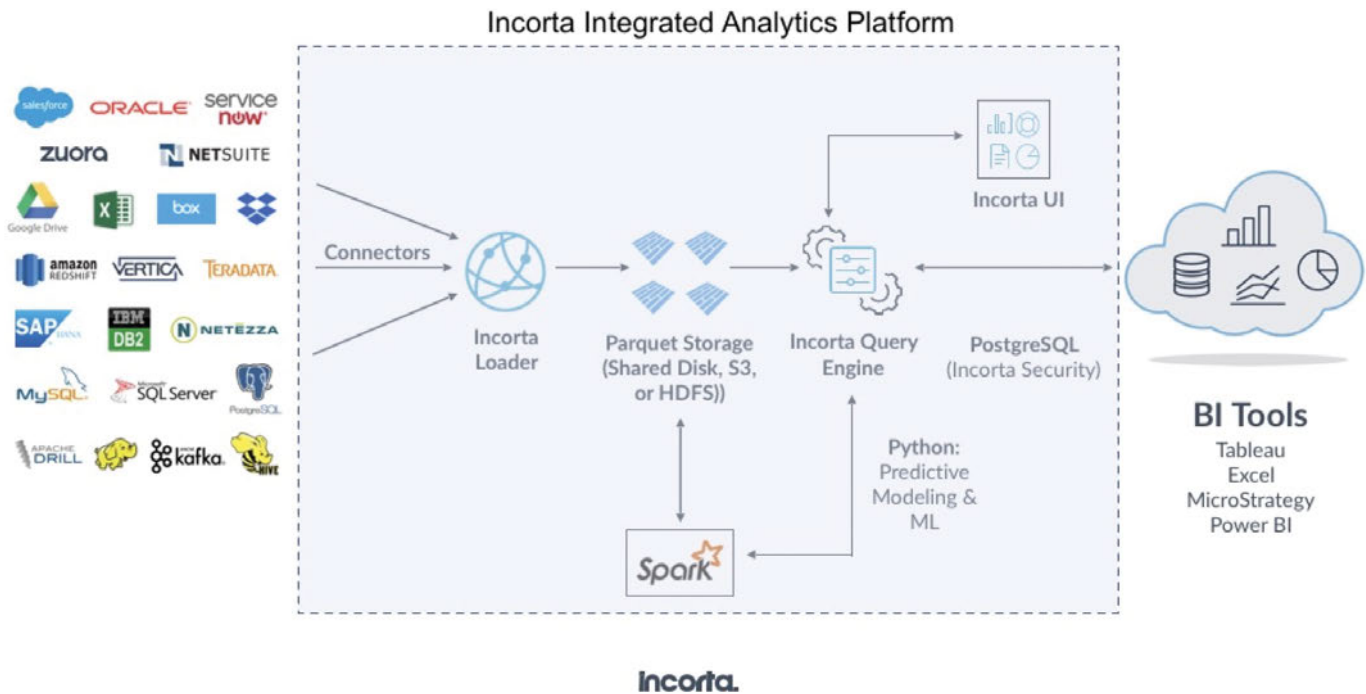
Finally, undertake new data warehousing and analytics work as Direct Data Mapping™ projects that produce results quickly, with low latency and high adaptability. By migrating legacy systems and building for new requirements with common technology, the path to modern data warehousing is clear and achievable.

Despite many premature declarations of the death of data warehousing, we will continue to need integrated, subject-oriented, non-volatile, and time-variant data for the foreseeable future. We need data warehousing capabilities without the burdens of the traditional data warehouse. Incorta's Direct Data Mapping makes it all possible.



Need help with your business analytics or data management and governance strategy?
Want to learn about the latest business analytics and big data tools and trends?
Check out **Eckerson Group** research and consulting services.

About Incorta



Incorta aggregates large, complex business data in real-time, eliminating the need to reshape your data into analytical formats. Utilizing the industry’s first Direct Data Mapping engine, Incorta delivers unprecedented join performance making the data warehouse obsolete. By keeping data in its original form, Incorta accelerates the time required to roll out new analytic applications from months to days, and reduces query and reporting times from minutes to seconds. With Google-like search and direct integration with Excel, Splasht BI and other visualization tools. Incorta allows business users to easily, securely tap into their enterprise data.

To learn more, visit www.solutions4less.tech or call us at 719-494-6182. This report is a companion to the webinar, [Modern Data Warehousing without the Burden of ETL.](#)